# EMV: Why Payment Systems Fail

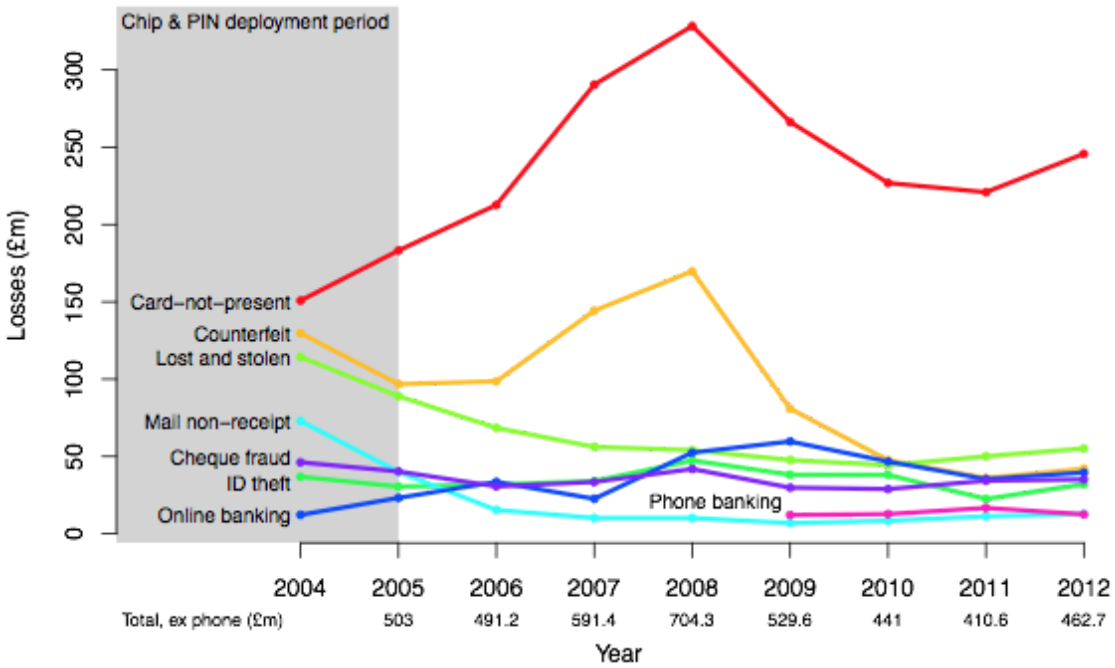
Ross Anderson and Steven Murdoch

## Abstract

What lessons might we learn from the chip cards used for payments in Europe, now that America's adopting them too?

## Introduction

Americans are starting to get new credit cards with an embedded chip as well as the magnetic strip that has been in use since the 1970s. Named for its promoters Europay, MasterCard and Visa, the EMV system augments the old magnetic strip cards with a chip that can authenticate a transaction using cryptography – a so called "smart card". EMV was deployed in the UK from 2003–6 and in other European countries shortly afterwards; it's now been rolled out from India to Canada. The idea was to cut fraud drastically; but reality turned out to be somewhat harder than theory. As the graph shows, fraud in the UK went up, then down, and is now heading upwards again.




The idea behind EMV is simple enough. The card is authenticated by a chip that's a lot harder to forge than the magnetic strip. The cardholder may be identified by a signature as before, or by a PIN; the chip has the ability to verify the PIN locally. Banks in the UK decided to use PIN

verification wherever possible, so the system there is branded “chip and PIN”; in Singapore, it’s “chip and signature” as banks decided to continue using signatures at the point of sale. America looks like being a mixture, with some banks issuing chip-and-PIN cards and others going down the signature route. We may therefore be about to see a large natural experiment as to whether it’s better to authenticate transactions with a signature or a PIN.

The key question will be, “better for whom?” The European experience suggests that this will not be a straight fight between the fraudsters and everyone else. The interests of banks, merchants, regulators, vendors and consumers clash in interesting ways; the outcome won’t just be determined by how the fraudsters adapt to the technology, but by a complex tussle over who pays for the upgrade and who enjoys the benefits. Fraud savings are not the biggest game in town; while fraud costs America $3–4 billion, interchange fees are a whopping $30 billion and EMV will likely have an impact on both.

## Attacks

Although US banks are issuing EMV cards now, it will be some time before they start to see a reduction in fraud. Cards will still have the magnetic strip and banks will continue to accept magnetic strip transactions because it will take many years to upgrade all the ATMs and shop terminals. EMV terminals still process unencrypted card numbers, expiry dates and PINs, so if hacked (as occurred in the 2014 Target data breach), criminals can steal enough data to perform fraudulent online purchases. Also, as many chip cards still contain a full unencrypted copy of the magnetic strip data, the criminals can steal this. If they can get the PIN too, they can make forged cards and use them at an ATM.

EMV also introduces some new vulnerabilities. The first-wave EMV cards in the UK were cheap cards capable only of Static Data Authentication (SDA), where the card contains a certificate signed by the bank attesting the card data are genuine. Since this certificate is static, it is trivial to copy it to a counterfeit chip, which can be programmed to accept any PIN – a so-called “Yes-card”. Criminals exploited this flaw at a small scale in France, but elsewhere it was not a serious problem. The Yes-card attack can be defeated by online transactions where the merchant contacts the bank to verify that the card computed a correct message authentication code on the transaction data. (This uses a key shared between the card and the issuing bank, so the merchant must be online for the code to be checked.) More modern EMV cards also support Dynamic Data Authentication (DDA) which uses asymmetric cryptography and defeats the Yes-card attack even for offline transactions. It is likely that US-issued EMV cards will support DDA and the vast majority of transactions will be online anyway, so the Yes-card attack is unlikely to be a major issue in the USA.

A much more serious type of fraud in the UK was tampering with Chip and PIN terminals to record card details and customer PINs. Although terminals were certified to be tamper resistant, they weren’t, and the certification process was seriously flawed [2]. Criminals were able to

modify terminals on a large scale to collect customer details as the card sent them to the terminal, and the PIN entered by the customer as it was sent to the card for verification. Because most UK cards stored a copy of the magnetic strip on the chip, criminals could then make fake magnetic-strip ATM cards. Before the use of chip and PIN in the UK, customers signed for store transactions and PINs were only used at ATMs, so tampering with a store terminal didn't yield enough information to withdraw cash from an ATM. Chip and PIN changed that; as merchants started accepting PINs at the point of sale, card forgery became easier and more prevalent. As the graph shows, counterfeit card fraud went up after EMV was deployed, and took five years for it to fall back to the previous level. So US banks can expect a lot of attacks using compromised or counterfeit terminals until they can start turning off magnetic-strip fallback mode.

Another attack we worried about in the early days of EMV was the "relay attack". This exploits the fact that while the card authenticates itself to the merchant terminal, the customer doesn't know which terminal the card is communicating with. If a customer inserts her card into a fake terminal, it can relay a transaction with a quite different terminal. So a crook, Bob, can set up a fake parking meter in New York, and when an unwitting cardholder Alice uses it, Carol (who's colluding with Bob) can stroll into Dave's jewelry store in San Francisco and buy a diamond using a fake card connected to the reader in the parking meter. The poor cardholder thinks she paid $20 for a parking space, and gets a statement showing she spent $2,000 in a store she never visited. The counterfeit card inserted into the genuine terminal simply relays the transaction back to the genuine card via the fake terminal (see Figure 1). While there's no real defence against the relay attack, it doesn't scale well, so is likely only going to be used against high-value targets.

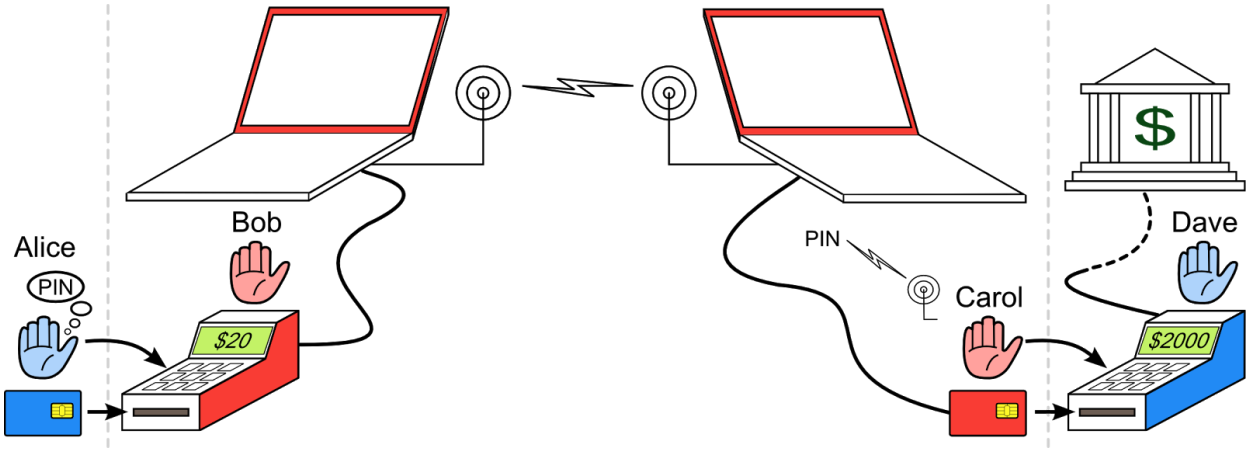



*Figure 1: The relay attack*

A more serious vulnerability is the No-PIN attack [1]. Here a criminal who has stolen a card but doesn't know the correct PIN can put a small electronic device between the stolen card and the terminal and use it with any PIN he likes. The device tricks the card into believing it's doing a chip and signature transaction while making the terminal believe that the card accepted the PIN

that was entered. This attack works against all types of card, and even for online transactions. Fixing it properly would require a change to the EMV protocol, which would take years to agree. In the meantime it is often possible for the card-issuing bank to detect the the attack by carefully comparing the card's version of the transaction with the terminal's. So far it appears that only one UK bank is trying to do so. In the meantime French criminals have been caught exploiting a more sophisticated variant of this attack in the wild.

The latest family of attacks, seen in the last two or three years in Spain, exploits a classic cryptographic vulnerability – the way in which EMV systems generate and use random numbers. When a terminal initiates an EMV transaction, it sends the card not just the date and the amount but a random number, so that each transaction is different and a crook cannot simply replay old transactions. However there are two flaws in this system. The first is an implementation flaw: it turns out that some ATMs generate predictable "random" numbers, so an attacker who has temporary access to someone else's card (say, a waiter in a Mafia-owned restaurant) can calculate an authentication code that he can use at some predictable time in the future at a known ATM. Worse, there is a design flaw in that the terminal does not transmit its choice of random number to the bank in an authenticated way. This means for example that if a terminal is running malicious software, it can harvest from a customer's card a series of authentication codes which it can then use to make extra transactions in the future, and it can fix up the random numbers in the protocol so that the issuing bank doesn't notice anything suspicious [3]. This is a serious attack because it can scale; a crime gang that managed to install malware on a number of legitimate terminals (as happened in the Target case) could harvest authentication codes to authorise large numbers of transactions at businesses under its control.

Finally, the elephant in the room with EMV deployment is card-not-present (CNP) fraud (Internet, phone and mail-order purchases). Although CNP fraud was low when EMV started to be deployed in the UK, it had grown to over half of UK card fraud by the time the roll-out was complete. EMV does almost nothing to stop CNP fraud (cards were never designed to be connected to customer PCs, even if smart card readers were to become prevalent) and so the crooks' initial reaction to the EMV deployment was just to take their business online, as we can see from the graph. US banks would be well advised to invest in further measures to mitigate CNP fraud rather than putting their entire security budget into deploying EMV. EMVCo (the consortium which maintains the EMV standard) has already started work on a "tokenization" specification, allowing CNP transactions to be performed with limited-use tokens (in effect, one-time credit-card numbers) rather than a static credit card number, so reducing the damage resulting from merchant data breaches or malware on customer PCs. Tokenization has almost nothing to do with EMV chips, but rather than setting up an new industry body, the banks have drafted in EMVCo to deal with it.

## The business battleground

When credit cards were first introduced by Diners' Club in the 1960s, they had high fees; typically the merchant paid the bank 6% or more of transactions. The emergence in the 1970s of the Visa-MasterCard duopoly stabilised things with standard contracts for banks and merchants, technical standards so their computers could swap data, and standard fees at 2.5% for credit cards and 1.5% for debit. This enabled a huge expansion of the industry, and cards became the standard way to pay for items costing more than a few dollars. Card transaction processing has become a huge money-spinner for the banking industry, especially as the clunky old addressograph machines for taking paper imprints were replaced by cheaper online systems, and as card payments spread online too. Many merchants see the card industry as an exploitative cartel, in need of trustbusting or competitive innovation. In 2005 merchants filed a class-action suit against Visa and Mastercard; a settlement in 2013 lowered fees by 0.1% and allowed merchants to charge customers the higher costs of credit-card transactions (which they already do in Europe). There has also been legislative action, with the Durbin amendment to the Dodd-Frank bill empowering the Federal Reserve to write the rules for fees on debit card transactions.

The sums involved are large. A retailer like Walmart, for example, takes over $200bn in credit-card sales; if these customers could be moved to PIN-based debit card transactions, that would save $2bn in fees. So some retailers have strongly supported the move to EMV. At the same time, the versions of the EMV protocol being introduced in the US to support contactless payments (such as where your credit card becomes an app on your NFC mobile phone) are designed to make it harder for merchants to move customers to debit card payments. These market dynamics are unlike those seen in Canada or Europe, where the banking industry motivated merchants to install EMV terminals by means of a "liability shift": the banks changed their terms and conditions so that merchants were charged the cost of all customer disputes where a PIN was not used. Where a PIN was used, the banks would then pass the liability on to the cardholder, saying "Your chip was read and your PIN was used, so you must have been negligent or complicit." Such a liability shift would be more difficult in the USA because the retailers' lobby is as powerful as the banks, and because consumer protection is better entrenched in US regulation.

Yet consumer protection may be undermined in a multitude of ways. One example is the protocol used to decide how to authenticate the cardholder. According to the EMV standards, each card has a cardholder verification method (CVM) list which states a preference such as 'first, signature; then PIN'; the terminal should read this and use the highest-ranked method it supports. We would expect to see aggressive retailers programming their terminals to insist on a PIN whenever that's supported, if (as we expect) the fee or liability for a PIN-based transaction is lower. In fact we have come across cases where merchants have simply lied to the banks about the method used. One fraud victim whose card was stolen while he was on holiday in Turkey was denied a refund for a charge made to his card because the merchant reported it as PIN-based; he managed to get a copy of the till receipt and found that the thief had in fact signed for the goods. If you wish to avoid this sort of problem, it is prudent to demand a card that only supports chip-and-signature.

Indeed, as the US will be the first country with a mixture of chip-and-pin and chip-and-signature cards in issue, we should be able to learn a lot from the crime figures after a few years. And this is not just about fraud, but robbery too. Chip-and-PIN cards are typically capable of offline PIN verification, and European banks have issued millions of card readers which enable cardholders to compute authentication codes for online banking. These readers can be used by muggers to check whether a victim is telling the truth when they demand his PIN as well as his cards; in one unfortunate case, two French students were tortured to death by robbers.

The most widespread problem encountered by cardholders, though is likely to be in dispute resolution. One of the problems thrown up by the experience in Europe is the lack of suitable tools for courts, arbitrators and even front-line dispute resolution staff in banks. When disputes arose with magnetic-strip cards, the consumer typically got the benefit of the doubt as these were widely known to be forgeable. EMV systems on the other hand create large amounts of log data which appear to be impressive but are often not understood, and can sometimes be the result of forgery by merchants (as in the Turkish case) or by malware on merchant terminals (as in the recent Target case, which would likely have been unaffected by the move to EMV). Also, the move from signature to PIN verification shifted dispute resolution in the banks favour. Any forged signature will likely be shown to be a forgery by later expert examination. In contrast, if the correct PIN was entered the fraud victim is left in the impossible position of having to prove that he did not negligently disclose it.

The main lesson to be learned here is that the collection, analysis and presentation of evidence is a function that needs to be specified, tested and debugged like any other. Simply dumping many pages of printout on a court and leaving it to an expert to pore through the digits, comparing them with EMV manuals, is not a robust way to do things; often the necessary evidence isn't even retained. The forensic procedures also need to be open and transparent to stand up in court, and their governance needs to be improved; this problem cannot just be left to a disparate vendor community [5]. Here some guidance from the Fed would be welcome.

## Conclusions

The EMV protocol is not a rigid way of doing card payments so much as a toolkit with which banks can build systems that can be pretty secure, but which can also be pretty awful. There is good news, and bad news. The good news is that EMV systems have been deployed in Europe for eleven years now, and there is a lot of experience to build on. Almost everything that could go wrong, has gone wrong: several protocol flaws which allowed attacks nobody had anticipated; tamper-resistance that didn't work; certification schemes that turned out to be a sham; and evidence-collection systems that were not fit for purpose. These should not just be academic case studies for security engineering classes, but should be studied by engineers who want to build robust payment systems.

The bad news is that the interests of banks, merchants, vendors, cardholders and regulators diverge in significant ways. In Europe, many failures were down to banks dumping liability on merchants and cardholders, who were in no position to defend themselves. In the US, the dynamic is different and more complex, with the main fight being over the interchange fees that the merchants pay the banks for processing their transactions. These fees are an order of magnitude greater than the fraud is, so we may find that the security of the system will be a side-effect of the project rather than its main goal. The details may be fought over for years to come in the courts and by lobbyists in Washington.

[1] Chip and PIN is Broken, Steven J. Murdoch, Saar Drimer, Ross Anderson, Mike Bond. IEEE Symposium on Security and Privacy, Oakland, CA, US, 16–19 May 2010.
[2] Thinking Inside the Box: System-level Failures of Tamper Proofing, Saar Drimer, Steven J. Murdoch, Ross Anderson. IEEE Symposium on Security and Privacy, Oakland, CA, US, 18–21 May 2008.
[3] Chip and Skim: cloning EMV cards with the pre-play attack, Mike Bond, Omar Choudary, Steven J. Murdoch, Sergei Skorobogatov, Ross Anderson. IEEE Symposium on Security and Privacy, San Jose, CA, US, 18–21 May 2014.
[4] Keep Your Enemies Close: Distance Bounding Against Smartcard Relay Attacks, Saar Drimer, Steven J. Murdoch. USENIX Security Symposium, Boston, MA, USA, 06–10 August 2007.
[5] Security Protocols and Evidence: Where Many Payment Systems Fail, Ross Anderson, Steven J. Murdoch. Financial Cryptography and Data Security, Barbados, 03–07 March 2014.

*(version accepted by Communications of the ACM, March 2014; appeared June)*